An Intuitive Guide to pdf-Transformation (for non-mathematicians)

Min Xu

September 20, 2010

If $(X_1, ... X_n)$ is a random vector whose range is $\mathbb{R}^n$, and $(Y_1, ... Y_n) = (\phi_1(X_1, ... X_n), ... \phi_n(X_1, ... X_n))$ such that for all i, j, $\frac{\partial \phi_j}{\partial x_i}$ exist and there exist an inverse functions $\phi_j^{-1}(Y_1, ... Y_n) = X_j$, then

$$f_Y(y_1, ... y_n) = f_X(\phi_1^{-1}(y_1, ... y_n) ... \phi_n^{-1}(y_1, ... y_n)) \begin{vmatrix} \frac{\partial \phi_1^{-1}}{\partial y_1} & ... & \frac{\partial \phi_1^{-1}}{\partial y_n} \\ ... & & \\ \frac{\partial \phi_n^{-1}}{\partial y_1} & ... & \frac{\partial \phi_n^{-1}}{\partial y_n} \end{vmatrix}$$

Where we take absolute value of determinant of the Jacobian matrix.

The goal of this guide is to give the intuition behind this formula.

1 One Variable Case

First, say X is a random variable instead of a random vector and $Y = \phi(X)$. Assume also that ϕ is differentiable and ϕ^{-1} exist (ϕ^{-1} would also be differentiabe).

Now, if $\Delta > 0$ is very small, then by definition of probability density function

$$f_Y(y_0)2\Delta \approx P(Y \in (y_0 - \Delta, y_0 + \Delta))$$

Where 2Δ is a positive number representing the length of the small interval.

As Δ get smaller and smaller, the $\approx$ becomes more and more accurate. In the limit as $\Delta \to 0$, the $\approx$ becomes $=$, so we'll just use $=$ in its place from now on.

This implies

$$f_Y(y_0)2\Delta = P(X \in (\phi^{-1}(y_0 - \Delta), \phi^{-1}(y_0 + \Delta)))$$

We know based on the definition of derivative:

$$\frac{d\phi^{-1}(y_0)}{dy} = \lim_{\Delta \to 0} \frac{\phi^{-1}(y_0 + \Delta) - \phi^{-1}(y_0)}{\Delta}$$

That if Δ is really small, then

$$\phi^{-1}(y_0 + \Delta) = \phi^{-1}(y_0) + \Delta \frac{d\phi^{-1}(y_0)}{dy}$$

Hence, going back to probability,

$$f_Y(y_0)2\Delta = P(X \in (\phi^{-1}(y_0 - \Delta), \phi^{-1}(y_0 + \Delta))) \tag{1.1}$$

$$= P(X \in (\phi^{-1}(y_0) - \Delta \frac{d\phi^{-1}(y_0)}{dy}, \phi^{-1}(y_0) + \Delta \frac{d\phi^{-1}(y_0)}{dy})) \tag{1.2}$$

$$= P(X \in (\phi^{-1}(y_0) - \Delta|\frac{d\phi^{-1}(y_0)}{dy}|, \phi^{-1}(y_0) + \Delta|\frac{d\phi^{-1}(y_0)}{dy}|)) \tag{1.3}$$

We put the absolute value because we want $\Delta|\frac{d\phi^{-1}(y_0)}{dy}|$ term to be positive to represent the length of the interval.

Now, using the definition of $f_X(x)$, we get that

$$f_Y(y_0)2\Delta = f_X(\phi^{-1}(y_0))2|\frac{d\phi^{-1}(y_0)}{dy}|\Delta$$

Canceling out 2Δ gives us

$$f_Y(y_0) = f_X(\phi^{-1}(y_0))|\frac{d\phi^{-1}(y_0)}{dy}|$$

So what happened is that ϕ stretched the Δ interval and we need the $|\frac{d\phi^{-1}(y_0)}{dy}|$ to "adjust" for the stretching.

2 Multi-variate Case

Now, suppose $X = (X_1, ...X_n)$ is a random vector, and $Y = (Y_1, ...Y_n) = (\phi_1(X_1, ...X_n), ...\phi_n(X_1, ...X_n))$.

Keeping in mind that y_0 is now a vector:

$$f_Y(y_0) \cdot (2\Delta)^n \approx P(Y \in Cube_\Delta(y_0)) \tag{2.1}$$

$$= P(X \in \phi^{-1}(Cube_\Delta(y_0)))) \tag{2.2}$$

Where $Cube_\Delta(y_0)$ is an n-dimensional cube that has y_0 at the center with side length 2Δ and $(2\Delta)^n$ is the volume of this cube.

Again, the approximation gets better and better as the volume of the Cube gets smaller and smaller. In the limit that the volume goes to 0, the approximation is exact.

Note that we didn't have to use a cube here, we could have used any solid so long as (1) we multiply $f_Y(y_0)$ by the volume of the solid and (2) y_0 is inside the solid. This will be important later.

Now, the Jacobian matrix $J(y_0)$ of ϕ^-1 is

$$\begin{bmatrix} \frac{\partial\phi_1^{-1}(y_0)}{\partial y_1} & ... & \frac{\partial\phi_1^{-1}(y_0)}{\partial y_n} \\ ... & & \\ \frac{\partial\phi_n^{-1}(y_0)}{\partial y_1} & ... & \frac{\partial\phi_n^{-1}(y_0)}{\partial y_n} \end{bmatrix}$$

And it has the property that

$$\lim_{||v|| \to 0} \frac{||\phi^{-1}(y_0 + v) - \phi^{-1}(y_0) - J(y_0)v||}{||v||} = 0$$

Where v is a vector, y_0 is a vector, and $\phi^{-1}(y_0)$ is a vector, and $J(y_0)$ is the Jacobian matrix.

Intuitively, this property is saying $J(y_0)v + \phi^{-1}(y_0)$ is a good estimate of $\phi^{-1}(y_0 + v)$ so long as v is a very small vector. This is analogous to the one variable case, where $\frac{d\phi^{-1}(y_0)}{dy}\Delta + \phi^{-1}(y_0)$ is a good estimate of $\phi^{-1}(y_0 + \Delta)$.

Now, imagine $Cube_\Delta(y_0)$ as a collection of really small vectors v's originating from the point y_0. If we apply ϕ^{-1} on the cube, that is the same as first translating the center from y_0 to $\phi^{-1}(y_0)$ and then applying matrix $J(y_0)$ on each of the these small vector v's to get a new n-dimensional solid centered on $\phi^{-1}(y_0)$. Let us call that new solid $Solid_J(\phi^{-1}(y_0))$.

You might remember that the determinant of a matrix has something to do with volume. To be precise, if v is a vector from origin, then you can think of v and the origin as forming the vertices of a rectangular solid. (For example, in three dimensional case, $v = (x, y, z)$ and $(0, 0, 0)$ forms the vertices a rectangular solid with side lengths x, y, x)

If you have matrix M, then you can change the vector v into vector Mv and thus change the Solid v into the Solid Mv. And it turns out that $|\det(M)|$ is exactly $\frac{Volume(Mv)}{Volume(v)}$. That is, the absolute value of determinant is the ratio of change in volume when a solid is changed by a matrix. (Google "Determinant and Volume" for more info)

Intuitively then, the volume of $Solid_J((\phi^{-1}(y_0))$ is $|\det(J(y_0))| \cdot (2\Delta)^n$ since $(2\Delta)^n$ is the volume of the old solid.

Therefore,

$$f_Y(y_0)(2\Delta)^n = P(X \in \phi^{-1}(Cube_\Delta(y_0)))) \tag{2.3}$$

$$= P(X \in Solid_J(\phi^{-1}(y_0))) \tag{2.4}$$

$$= f_X(\phi^{-1}(y_0)) \cdot Volume(Solid_J(\phi^{-1}(y_0))) \tag{2.5}$$

$$= f_X(\phi^{-1}(y_0)) \cdot |\det(J(y_0))|(2\Delta)^n \tag{2.6}$$

Canceling out the $(2\Delta)^n$ from both side, we get

$$f_Y(y_0) = f_X(\phi^{-1}(y_0))|\det(J(y_0))|$$

where $J(y_0)$ is the Jacobian matrix of ϕ^{-1} at y_0.

Thus, we see that when we apply ϕ, the Solid that contains y_0 get warped and changes volumes. The $|\det(J(y_0))|$ term is used to "adjust" for this volume change.